

not sent  
but note calculations.

April 4, 1948.

Dear Aaron,

Visits are always too short, as was your's, and will be ours. Following your suggestion, we plan to leave Thursday afternoon, April 15, and expect to arrive in Chicago via Milwaukee at 520PM. We have reservations for the Pacemaker which leaves at 330 PM on Friday. I think that you already understand these dates and times.

Hershey's paper has of course come out in PNAS, but I haven't really been able to study it very closely. The suggestion that the *r* mutants may be complex is still not negated, and I can't understand the relevancy of A.D.'s argument that the sequence *r* - *hr* - *r* was effected to this point. On the other hand, the approximate equivalence of "complex *r*'s" and + in crosses suggests a unifactorial basis. I wonder whether the spontaneous instability of the *r* character does not interfere with the determination of at least some of the more closely linked *r*'s. I wonder whether you should not formulate your "complex genotype" theory as subsuming two sets of factors of which one each must be mutant to give the *r* phenotype. By the way, I don't get nearly such high estimates of the total number of equimutable loci that would probably give no recurrences in a sample of 20. If *n* is the total number, the distribution of occurrences of a given mutant in samples of 20 will be given by  $p_r = e^{-20/n} (1, 20/n, \dots)$  in the Poisson series, and for *r* = 0 or 1 we have  $p_{-2} = e^{-20/n} (1 + 20/n)$ . For all *n* loci we have then  $p_{-2} = e^{-20} (1 + 20/n)^n$ . The following table gives  $\log p$  for a few values of *n*:

| <i>n</i> | $\log p$ | <i>p</i> |   |
|----------|----------|----------|---|
| 20       | -2.62    | .0024    | He could, then, have only 50 loci without too much improbability. |
| 40       | -1.64    | .023     |   |
| 50       | -1.38    | .042     | But there is no point pursuing                                    |
| 100      | -0.76    | .19      | this, since even twenty factors                                   |
| 200      | -0.40    | .40      | are somewhat incredible.  |
| 500      | -0.18    | .66      |   |

Using two sets of factors on your hypothesis has far more genetic precedent than the assumption that any two factors comutant lead to the *r* phenotypes, since it would ascribe qualitatively distinct functions to the different loci. However, this may not be the time to bring in antique ideas.

Have you enquired at all about galactosides, (also lactosides, a-l-arabinosides, and B-D-fucosides)?

See you anon,